



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
[www.uspto.gov](http://www.uspto.gov)

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/775,110	02/11/2004	Shigeru Tago	HIRA.0143	3621
7590	11/29/2006		EXAMINER	
Reed Smith LLP Suite 1400 3110 Fairview Park Drive Falls Church, VA 22042-4503				SKIBINSKY, ANNA
			ART UNIT	PAPER NUMBER
			1631	

DATE MAILED: 11/29/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

<b>Office Action Summary</b>	Application No.	Applicant(s)
	10/775,110	TAGO ET AL.
	Examiner Anna Skibinsky	Art Unit 1631

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

#### Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
  - If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
  - Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

#### Status

- 1) Responsive to communication(s) filed on 05 September 2006.  
 2a) This action is FINAL.                    2b) This action is non-final.  
 3) Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

#### Disposition of Claims

- 4) Claim(s) 1,3,4 and 6-8 is/are pending in the application.  
 4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.  
 5) Claim(s) \_\_\_\_\_ is/are allowed.  
 6) Claim(s) 1,3,4 and 6-8 is/are rejected.  
 7) Claim(s) \_\_\_\_\_ is/are objected to.  
 8) Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

#### Application Papers

- 9) The specification is objected to by the Examiner.  
 10) The drawing(s) filed on \_\_\_\_\_ is/are: a) accepted or b) objected to by the Examiner.  
     Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).  
     Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).  
 11) The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

#### Priority under 35 U.S.C. § 119

- 12) Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).  
 a) All    b) Some \* c) None of:  
     1. Certified copies of the priority documents have been received.  
     2. Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.  
     3. Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

\* See the attached detailed Office action for a list of the certified copies not received.

#### Attachment(s)

- |  |   |
|--|---|
| 1) <input type="checkbox"/> Notice of References Cited (PTO-892)   | 4) <input type="checkbox"/> Interview Summary (PTO-413)           |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948)   | Paper No(s)/Mail Date. _____                                      |
| 3) <input checked="" type="checkbox"/> Information Disclosure Statement(s) (PTO/SB/08)<br>Paper No(s)/Mail Date <u>9/05/06</u> . | 5) <input type="checkbox"/> Notice of Informal Patent Application |
|  | 6) <input type="checkbox"/> Other: _____                          |

## **DETAILED ACTION**

### ***Reply to Applicant's Amendments***

Applicant's amendments to claims 1, 3, 4, and 6 are acknowledged. The limitations of claims 2 and 5 have been amended into claim 1, lines 20-25 and claim 4, lines 23-26, and claims 2 and 5 have been cancelled.

### ***Objection to the Specification***

1. The disclosure is objected to because it contains an embedded hyperlink and/or other form of browser-executable code on page 1. Applicant is required to delete the embedded hyperlink (i.e. "http://www.") and/or other form of browser-executable code. See MPEP § 608.01.

### ***Objection to Claims***

2. Claims 6 and 8 are objected to because of the following informalities: Said claims recite "A program ..." but depend from claims 4 and 5 and should recite "The program ..." Appropriate correction is required.

### ***Claim Rejections - 35 USC § 112-1<sup>st</sup> paragraph***

The rejection of claim(s) 1-8 for New Matter under 35 USC § 112-1<sup>st</sup> paragraph in the Office Action filed September 5, 2006 is withdrawn in view of Applicant's Remarks/Amendments filed June 5, 2006

***Claim Rejections - 35 USC § 102***

The rejection of claim(s) 1, 3-4, and 6-8 over Murray et al. in the Office Action filed June 5, 2006 is withdrawn in view of Applicant's Remarks/Amendments filed September 5, 2006

***Claim Rejections - 35 USC § 103***

1. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

This application currently names joint inventors. In considering patentability of the claims under 35 U.S.C. 103(a), the examiner presumes that the subject matter of the various claims was commonly owned at the time any inventions covered therein were made absent any evidence to the contrary. Applicant is advised of the obligation under 37 CFR 1.56 to point out the inventor and invention dates of each claim that was not commonly owned at the time a later invention was made in order for the examiner to consider the applicability of 35 U.S.C. 103(c) and potential 35 U.S.C. 102(e), (f) or (g) prior art under 35 U.S.C. 103(a).

Art Unit: 1631

Claims 1, 3-4, and 6-8 are rejected under 35 U.S.C. 103(a) as being unpatentable over Murray et al (U.S. Patent No. 6,876,930) in view of Getchius (U.S. Patent No. 6,519,592).

Applicant claims a method of extracting data from two databases and calculating the frequency of appearance of keywords from a key word table. A tree structure is used to organize the keywords in a table.

Murray et al. teach extracting data from literature databases such as MEDLINE, related to a certain gene (col. 2 line 63 to col. 3 line 6) using a computer and programmed algorithms. The limitations recited in claim 1 are further described in col. 26, lines 13-50, col. 27, line 58 (Fig. 10) to col. 29, as well as throughout the body of the prior art. Murray et al. teach how literature relevant to genes in a list of genes is identified, downloaded and stored (col. 26, lines 14-15). Relationships between genes are then extracted from the literature using information extraction techniques, cross referenced and graphed (col. 26, lines 16-20). The data produced and graphed is stored in a 2<sup>nd</sup> database referred to as the "expert database" (col. 26, lines 20-26), as in instant claim 1, lines 9-10. Here the identifier (instant claim 1, lines 8-10) is the criteria used to extract the relationship between the genes. The data from the second database may then be viewed by the user who searches to verify gene expression results (claim 1, lines 11-12).

Furthermore, finding the frequency of a biological topic described in abstracts is taught. These published references are then scored based on the times they are referenced in other journals (col. 17, lines 34-47, col. 18, lines 1-4, and Table 1). The

prior art teaches the limitations of claim 1, lines 13-15. Here keywords can be the journal titles stored in a database along with the abstracts and articles.

1. Applicant amended (claim set 3/05/06) claim 1 to recite "a base sequence or an amino acid sequence of a gene or protein of interest inputted by a user". The prior art teaches the facilitation identification of candidate genes from a plurality of DNA sequences (paragraph 0010, lines 1-3). Information from literature databases relating to particular set of DNA sequences is retrieved to facilitate its identification as a candidate gene (paragraph 0011). Thus, the extraction of the DNA sequences in the prior art also contains a base sequence or a gene or protein of interest inputted by a user.
2. Applicant has amended (claim set 3/05/06) claim 1, to recite "extracting an identifier identifying document data in said first text data from said extracted first text data which contains the base sequence of the amino acid sequence". The prior art teaches extracting information from literature where the extracted information pertains to a gene of interest (paragraph 0013, lines 7-11). Here, the identifier is the gene of interest which is present in the "identifier identifying document".
3. Applicant has amended (claim set 3/05/06) claim 1 to recite "a second text data extraction step for extracting second text data from said second database which contains said extracted identifier". The prior art teaches the extraction of "one of many literature databases" such as MEDLINE, USPTO, etc. (paragraph 0013, lines 1-7). The prior art also can access, extract, and cross-reference information from a literature database pertaining to a gene of a known expression pattern. Again, here, the identifier

of claim 1 is the gene and the multiple databases accessible by the invention of the prior art provide a first and second database.

4. Applicants have amended (claim set 3/05/06) claim 1 to recite that the appearance frequency calculation step is for the reading of “keywords of known functions or characteristics of genes or proteins from” the fist database and for “calculating a frequency of each of the keywords in the extracted second text data.” The prior art teaches that there is a need for comprehensive collections of entity-relationships amongst genes, functions, roles and pathways. Information or “keywords” are extracted from journal abstracts that describe certain characteristics about genes and information in the articles are ranked as being the “truth” or “false” after being scored (paragraph 0098, line 12 to paragraph 0100). The frequency score can be for the journal titles that contain the articles (Table 1) with the “truth” related to a certain gene. These journal references are stored in database tables (paragraph 0105, line 12 to paragraph 0107). Here, the appearance of words or information pertaining to a gene’s function (e.g. gene is an enzyme) is stored in a table (e.g. EnzymeRef). Each table is then scored based on the frequency or the amount of informational “keywords” it has pertaining to that gene’s function.

5. Applicants have amended (claim set 9/05/2006) claim 1, lines 16-17, to recite calculating the frequency of appearance of keywords “by counting a number of second documents containing said keywords.” The Murrey et al. teaches produced and graphed data stored in a 2<sup>nd</sup> database referred to as the “expert database” (col. 26, lines 20-26)

which is accessible by the user. Furthermore, a second step of extracting information regarding relationships between gene products is taught (col. 26, lines 16-18).

6. Applicants have amended (claim set 3/05/06) claim 1, lines 18-19, to recite "a display step for displaying a frequency of appearance of each of said keywords in a corresponding position in said keyword table". Murray et al. describes a table of characteristics (paragraph 0106 and Figure 7) where the score for the popularity of each topic is stored under FunctionScore, RoleScore, etc. The table in the form of a tree is displayed in Figure 7 and the score is the frequency that a particular characteristic of the gene has been added to from a database of articles.

As recited in claim 3, Murray et al. teach extracting data for a plurality of gene sequences (e.g. col. 3, lines 4-6 and col. 6, lines 41-46) throughout the specification and do not limit query strategies to one sequence at a time.

As in recited in claims 4 and 6, a database environment is described (col. 7, line 10 to col. 10, line 9) by Murray et al. where the above described method is carried out with a computer. In light of the above analysis, claim 4 is drawn to a program and which contains the same limitations in the body of the claim as as claim 1 is also rejected. Murray et al. teach a computational method with software to accesses a database (e.g. paragraphs 0050 to 0055).

7. Newly added (claim set 3/05/06) claims 7 and 8 recite "a frequency of each category in the keyword table is the sum of frequencies of lower level categories belonging to the category". Figure 7 shows categories within a main category as "Role" that contains a final category "RoleScores". The "RoleScores" contains all of the

scores from each of the possible functions (described on page 12, paragraph 0106, lines 4-11). Here, the "lower level categories" are the specific functions which are scored. The totality of the scores for each of the categories (e.g. development, endocytosis, etc.) are stored in the category called RoleScores.

8. Applicants have amended the limitations of claims 2 and 5 (claim set 3/05/06) into claims 1, lines 20-25, and 4, lines 23-26, respectively.

9. The "tree structure" recited in claims 1, lines 20-25, and 4, lines 23-26, and throughout the specification of the instant application have been interpreted to mean a computational data structure that has the form of a tree. This is a generic type of structure used in the computational arrangement and accessing of data. It is obvious to one of skill in the computational art that text data, keywords, and sequences related to any topic including genes can be organized in the form of a tree structure. A tree structure allows computational data to be organized, categorized, and accessed efficiently by a user. The art of Murray et al. relies on computational analysis and clustering of data in a database (Murray et al. paragraphs 0044-0045) where a result is a tree-like classification of gene characteristics (Figure 7). Thus, it would be obvious to use the tree structure database of Getchius who also relies on computational analysis and clustering of data since a tree structure is a generic computational data structure for organizing any form of computational data (Getchius et al. Abstract and col. 1, line 48 to 2 line 6). Getchius teaches that the tree data structure is used as a part of an on line query tool while Murray et al. also teach a web-based (URL) query tool (Murray et al. paragraphs 0052 and 0056). Furthermore, a tree structure does not depend on the type

of data stored therein but is a data structure especially useful for the organization and query of data. Thus, one of skill in the art would have a reasonable expectation of success at utilizing computational tree structures as described by Getchius to organize the data described by Murray et al.

Murray et al. teach extracting data from literature databases such as MEDLINE, related to a certain gene (col. 2 line 63 to col. 3 line 6) using a computer and programmed algorithms. The limitations recited in claim 1 are further described in col. 26, lines 13-50, col. 27, line 58 (Fig. 10) to col. 29, as well as throughout the body of the prior art. Murray et al. teach how literature relevant to genes in a list of genes is identified, downloaded and stored (col. 26, lines 14-15). Relationships between genes are then extracted from the literature using information extraction techniques, cross referenced and graphed (col. 26, lines 16-20). The data produced and graphed is stored in a 2<sup>nd</sup> database referred to as the "expert database" (col. 26, lines 20-26), as in instant claim 1, lines 11-12. Here the identifier (instant claim 1, lines 8-9) is the criteria used to extract the relationship between the genes. The data from the second database may then be viewed by the user who searches to verify gene expression results.

Furthermore, finding the frequency of a biological topic described in abstracts is taught. These published references are then scored based on the times they are referenced in other journals (col. 17, lines 34-47, col. 18, lines 1-4, and Table 1). The prior art teaches the limitations of claim 1, lines 13-17. Here keywords can be the journal titles stored in a database along with the abstracts and articles.

Murray et al. also teach “tagging” particular nouns and verbs related to a particular gene (col. 28, lines 24-26) and using decision trees to extract information from text (col. 29, lines 39-45), as in claim 1, lines 20-25, which uses a tree structure to organize keywords in categories. However, Murray et al. does not teach the use of tree structures.

As recited in claim 3, Murray et al. teach extracting data for a plurality of gene sequences (e.g. col. 3, lines 4-6 and col. 6, lines 41-46) throughout the specification and do not limit query strategies to one sequence at a time.

As in recited in claims 4, 5, and 6, a database environment is described (col. 7, line 10 to col. 10, line 9) by Murray et al. where the above described method is carried out with a computer. However, Murray et al. does not teach the use of tree structures in the database environment as recited in claim 5.

Murray et al. do not directly teach a keyword table with a tree structure as recited in claims 1, lines 20-25, and claim 5, lines 23-26. However, tree structures have long been known the computational arts as effective ways to arrange data, especially for queries. Getchius et al. teaches the use of a tree structure for tabulating computational data (col. 21, lines 5-6 and col. 33, lines 17-26). To build the table, the frequency of entries for each category is counted (col. 20, lines 38-44), thus the frequency of each key word that makes up a category is known.

It would have been *prima facie* obvious to one of ordinary skill in the art at the time the invention was made to have implemented tree structures as a commonly used form of programmable data structure in the database structures of Murray. One of skill

Art Unit: 1631

in the art would have been motivated to use tree structures, as they are considered to be an effective way for arranging data that will need to be retrieved, for example, in a query. One of skill in the art would have had the further motivation to implement the tree structures of Getchius in the methods and structures of Murray et al. as it allows for faster and more efficient access to information than data that is arranged in a linear table or list. One of skill in the art would have had a reasonable expectation of success at utilizing tree structures in the methods and structures of Murray, as they were well known in the art and do not require any special programming or hardware. Therefore, the invention as a whole would have been *prima facie* obvious, absent evidence to the contrary.

#### **REPLY TO ARGUMENTS**

10. Applicant's arguments filed 9/05/2006 have been fully considered but they are not persuasive.

11. Applicants argue (Remarks, page 7, last paragraph) that the cited references do not teach or suggest "a second document extraction step for extracting a second document from said second database which contains the same identifier (as contained in the said first document) of the base sequence or the amino acid sequence of a gene or protein or interest."

12. In response to applicant's argument that the references fail to show certain features of applicant's invention, it is noted that the features upon which applicant relies are not recited in the rejected claim(s). Although the claims are interpreted in light of

the specification, limitations from the specification are not read into the claims. See *In re Van Geuns*, 988 F.2d 1181, 26 USPQ2d 1057 (Fed. Cir. 1993).

13. Claim 1, lines 11-12, currently recite a “a second document extraction step for extracting a second document from said second database which contains said extracted identifier.” Applicant’s amendment of “document” to replace “text data” does not affect the scope of the claimed invention. As reiterated in the above rejection, Murrey et al. teaches produced and graphed data stored in a 2<sup>nd</sup> database referred to as the “expert database” (col. 26, lines 20-26) which is accessible by the user. Furthermore, a second step of extracting information regarding relationships between gene products is taught (col. 26, lines 16-18).

14. Applicants have amended claim 1 to recite “counting a number of second documents containing said keywords,” which is taught by Murray et al wherein the data extracted in the second step is extracted and then cross referenced which suggests that it is counted (i.e. accounted for).

15. Applicant argues (Remarks, page 8, lines 3-6) that Murray et al. does not involve any “base sequence or an amino acid sequence” of a gene or protein of interest.

16. In response, Murray et al. does teach involvement of amino acid sequences in that the invention facilitates identification of candidate genes from a plurality of DNA sequences (col. 2, lines 26-34).

17. Applicants argue (Remarks, page 8, lines 6-9) that Murray et al. does not “link genes/proteins of known functions/characteristics ...”

18. In response to applicant's argument that the references fail to show certain features of applicant's invention (i.e. "link[ing] genes/proteins of known functions/characteristics ...") , it is noted that the features upon which applicant relies are not recited in the rejected claim(s). Although the claims are interpreted in light of the specification, limitations from the specification are not read into the claims. See *In re Van Geuns*, 988 F.2d 1181, 26 USPQ2d 1057 (Fed. Cir. 1993).
19. Applicant's argue (Remarks, page 8, last line to page 9, lines 1-2) that Murray et al. does not teach "automatically and mechanically counting per document (not involving human opinion)" which amounts to the adjusted quantity scores of Murray et al. being different than that of the instant application.
20. In response to applicant's argument that the references fail to show certain features of applicant's invention, it is noted that the features upon which applicant relies are not recited in the rejected claim(s). Although the claims are interpreted in light of the specification, limitations from the specification are not read into the claims. See *In re Van Geuns*, 988 F.2d 1181, 26 USPQ2d 1057 (Fed. Cir. 1993).
21. Applicant's argue that Getchius et al., upon which Examiner relied for the teach of tree structures for the keyword table as recited in now cancelled claims 2 and 5 (currently amended into claim 1, lines 20-25 and claim 4, lines 23-26), is applied to a business listing rather than any "known functions or characteristics of genes or proteins."
22. In response, the key word table having a tree structure is a computational data structure used in the arts of computer science to which the instant application also

applies. As reiterated in the above rejection, both Getchius et al. and Murray et al. teach accessing information stored in database. Arranging data as a tree structure allows for faster and more efficient access to the information. Any generic computational data is benefited by a tree structure arrangement, as known in computer science art.

### ***Conclusion***

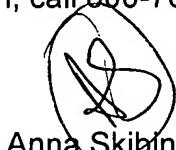
**THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

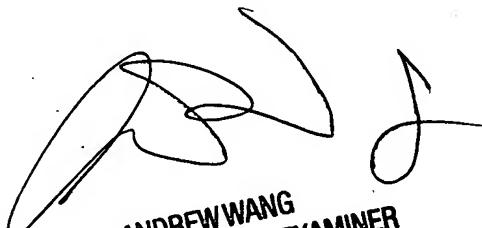
Any inquiry concerning this communication or earlier communications from the examiner should be directed to Anna Skibinsky whose telephone number is (571) 272-4373. The examiner can normally be reached on 8 am - 5:30 pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Andrew Wang can be reached on (571) 272-0811. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.



Anna Skibinsky, PhD



ANDREW WANG  
SUPERVISORY PATENT EXAMINER  
TECHNOLOGY CENTER 1600